

The Acoustic Theory of Vowel Nasalization

Handout by Julian Vargo
Linguistics 210: Advanced Phonetics
November 25, 2025

Overview

- Coarticulatory nasalization is a highly variable process that is difficult to measure without airflow data or some kind of articulatory data
- The literature on nasal vowels is quite 'stuffed up' (sorry)
 - There is no consensus about which acoustic cues are best for predicting vowel nasality. A few main cues stick out as useful (A1-P0 and F1 Bandwidth), but results differ wildly from study to study.
- In this lecture, I'll review the current methods in the field for measuring nasalization, and we'll review real-world data for interpreting nasalization with non-invasive methods

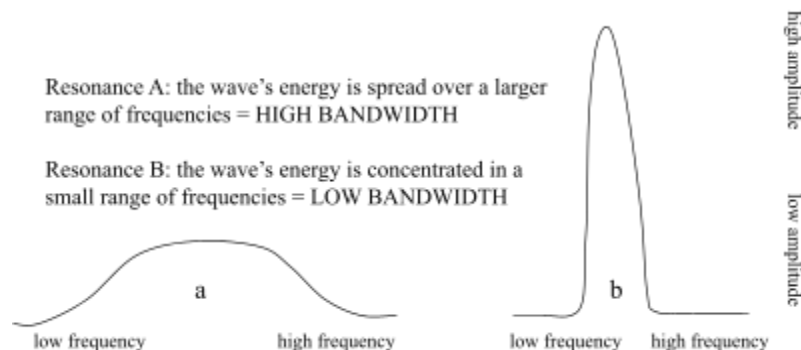
Key Terms & Concepts:

Maxillary sinus A large cavity which is connected to your nose. The larger the maxillary sinus, the more surface area is in your vocal tract (Sundberg et al. 2007).



VPO Velopharyngeal opening - the space between your nasal and oral cavity. Also called the *velopharyngeal port* or *velopharyngeal sphincter*.

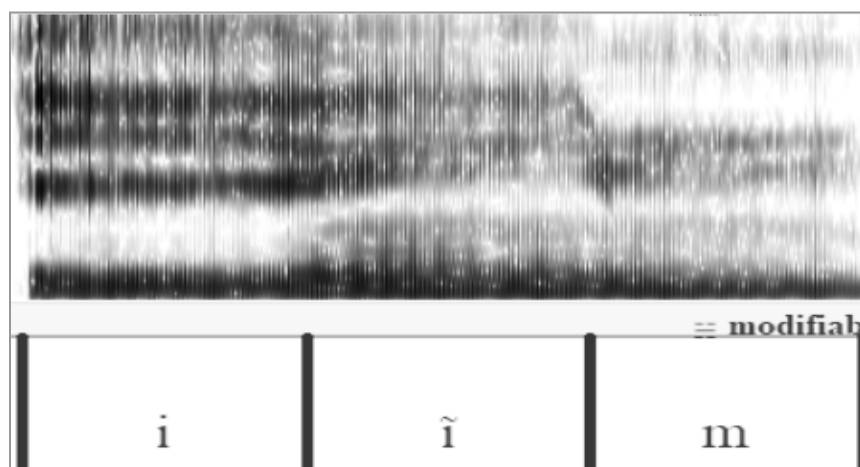
Formant bandwidth The width of the formant. This is a measure of how spread a resonance is.



Nasal sounds give the vocal tract a larger surface area because now the sound waves are bouncing around your nose and sinuses. The nose/sinus tissue absorbs more of the sound and disperses it over a wider bandwidth.

Resonance A = high bandwidth = nasal sound = open velopharyngeal port
Resonance B = low bandwidth = oral sound = closed velopharyngeal port

A1-P0	The amplitude of the first formant minus the amplitude of the first nasal 'pole' (also known as a nasal formant)
A1-P1	The amplitude of the first formant minus the amplitude of the second nasal 'pole' (also known as a nasal formant)
Antiformant	A band of low-amplitude energy arising from destructive interference from a side-resonator (nasal cavity, lateral side pockets, trachea). This looks like an extra white region of a spectrogram.



*Above: A highly exaggerated recording I made.
Note the fairly visible nasal poles and zeros*

1) The degree of vowel nasalization differs by language, by social factors

- English: widespread anticipatory vowel nasalization
 - 'cat' [k^hæʔ] ~ can't [k^hæʔ]
- English: nasalization can be used in stylistic performances and to build stereotyped personae (Podesva 2013):
 - high nasalization = 'nerd', 'librarian', 'redneck', 'party girl', 'rich snob'
 - low nasalization = 'bro', 'gangsta', 'cougar', 'mean old lady', 'sassy gay guy'
- Spanish: not widely described in most varieties
 - However, varieties with [ŋ] have more nasalization (Caravedo 1990, Bongiovanni 2020, Bongiovanni 2021, Campos-Astorkiza 2012:99)

2) The degree of vowel nasalization differs by speaker, by day

- What if my nose is bigger/smaller than your nose?
- What if my nose is stuffy? What if I'm sick? What if it's dry outside?

3) We can never elicit truly casual speech in laboratory contexts

- A speaker in a lab will never speak casually
- We can never know how much vowel nasalization is present in casual speech with 100% accuracy

- What if all we have are archived recordings? What if your speakers are old or cannot come into the lab? What if your department can't afford a nasometer?
 - Laboratory contexts have significant drawbacks, so it's crucial that we build our acoustic understanding and improve non-invasive methods.

4) *Simple articulation* → *complicated acoustics*

- F1 is a proxy for height, F2 is a proxy for backness... What about nasals?
 - Antiformants, Nasal Formants, F1 Bandwidth, A1-P0, A1-P1, F3 Bandwidth, Spectrum Standard Deviation, with each weight differing by speaker
- Styler (2015 & 2017): "The list of promising features was thus narrowed to four: A1-P0, Vowel Duration, Spectral Tilt, and Formant Frequency/Bandwidth."
 - methods: Support Vector Machines (SVM), Random Forests
- Carignan (2021): A1-P0 is the best predictor, but many other predictors are relevant
 - methods: linear regression, PCA, GAMM, Monte Carlo Markov Chains
- Carignan et al. (2023): Generate a continuous nasalization probability using XGBoost
- Pruthi & Espy Wilson (2004 & 2007): There are many relevant cues! Number of harmonics and spectral tilt stand out from other work.
 - methods: Hidden Markov Models, SVM
- Sundberg et al. (2007): A small change in velum opening and sinus size can massively change the acoustic output.
 - methods: build an epoxy model of a vocal tract! Incredibly unique methodology!
- Vargo (2025, recent HLS presentation): Bandwidth is far more important for Spanish than English, and English alters phonation to perceptually boost nasalization
 - method: feature importance & binary classification with XGBoost

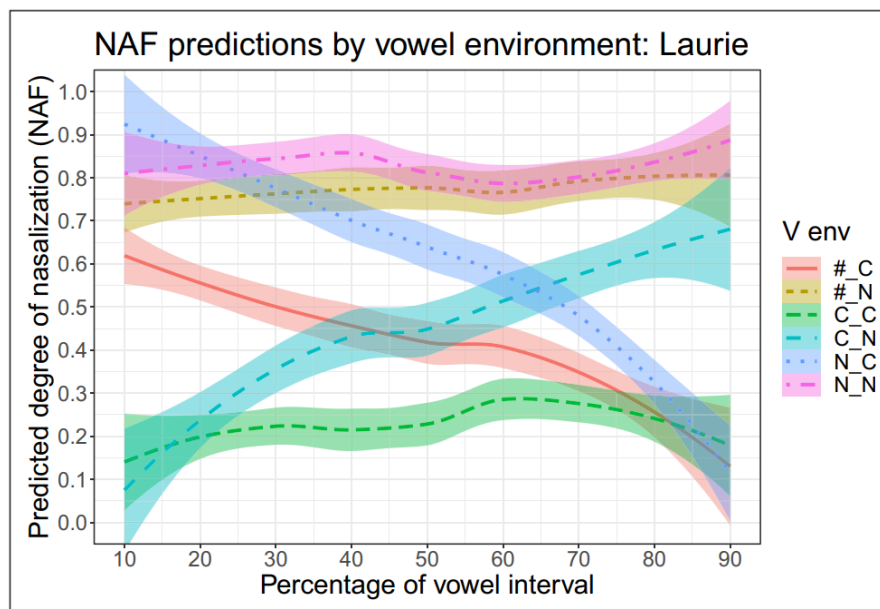
Lab: Replication of Carignan et al. (2023) in R

lab repo: <https://github.com/julian-vargo/nasalization-demo>

DOI: <https://doi.org/10.16995/labphon.9152>

- There are *many* tweaks that I make to the methodology, but we'll be keeping the same basic principles and goals in the spirit of the paper.
- Arabana is typologically unusual (at least for its Australian sprachbund) for contrasting /N/ with /TN/
- Research question: How does such a sound change arise? How can we measure nasalization to inform our theory of sound change?
- Hypothesis: two historical syllable structures
 - *NVNV
 - *CVNV
- A sound change happens where the nasal gesture drifts rightwards
 - Imagine a gestural score where your oral gestures are maintained but nasalization slips. This might remind you of Beddor (2009)
 - Duration(C) is relative to duration(N)
- *CVNV → CVTNV → VTNV

- $*NVNV \rightarrow N\tilde{V}NV \rightarrow \tilde{V}NV$
- The first nasal consonant blocks the sound change
- Once the initial consonants are deleted, we have an environment with bleeding and subsequent phonologization of /NT/ vs /N/
- May be a nasal/oral vowel contrast being phonologized?



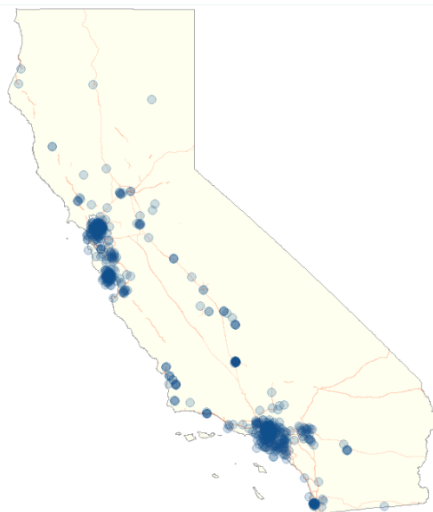
Above: The key figure from Carignan et al. (2023). Nasalization 'airflow' tracks can effectively be gathered through purely acoustic methods

- Train machine learning data:
 - 10% vowel timepoint of a CV sequence
 - 90% timepoint of a VC sequence
- Use a decision tree to measure whether a vowel is binarily nasal or oral
- Plot predicted probabilities on a continuous scale as GAMs



MuHSiC
MULTILINGUAL HISPANIC
SPEECH IN CALIFORNIA

- Multilingual Hispanic Speech in California Corpus (Amengual, Kim, Davidson & Vargo 2025)
 - 603 speakers
 - 35 minute interview in Spanish
 - 35 minute interview in English
 - 700+ hours of audio
 - 16 million phonemes



For our lab, we'll use Berkeley's own MuHSiC Corpus!